Trends in Cognitive Sciences

CellPress
REVIEWS

Review

Object Vision in a Structured World

Daniel Kaiser,[1,*] Genevieve L. Quek,[2] Radoslaw M. Cichy,[1,3,4] and Marius V. Peelen [2,*]

In natural vision, objects appear at typical locations, both with respect to visual space (e.g., an airplane in the upper part of a scene) and other objects (e.g., a lamp above a table). Recent studies have shown that object vision is strongly adapted to such positional regularities. In this review we synthesize these developments, highlighting that adaptations to positional regularities facilitate object detection and recognition, and sharpen the representations of objects in visual cortex. These effects are pervasive across various types of high-level content. We posit that adaptations to real-world structure collectively support optimal usage of limited cortical processing resources. Taking positional regularities into account will thus be essential for understanding efficient object vision in the real world.

Highlights

The characteristic spatial distribution of individual objects impacts profoundly on object processing in the human brain. Neural representations are sharper and perceptual sensitivity is increased when objects appear at retinal locations corresponding to their typical location in the world (e.g., an airplane in the upper visual field).

The characteristic positioning of objects relative to one another also systematically influences object processing. When multiobject displays are arranged in their typical relative positions (e.g., a lamp above a table), the visual system represents them as groups, allowing objects to be more easily detected, recognized, and memorized.

Neural adaptations to natural scene structure enable the visual brain to optimally represent the large number of objects contained in real-world environments, thereby simplifying the neural code for scene analysis.

Positional Regularities in Object Vision

Many natural behaviors crucially depend on accurately perceiving objects in the environment. Consequently, understanding object vision has been a core endeavor in cognitive neuroscience for many years, and recent decades have yielded exciting insights into how the human visual system processes various types of objects [1–5]. By and large, these insights have come from studies investigating the processing of individual objects presented at arbitrary locations (usually at fixation). However, in natural vision many objects often appear in specific locations both with respect to visual space (e.g., airplanes in the sky) and relative to other objects (e.g., lamps above tables).

Although it has already been well established that such real-world positional regularities furnish observers with cognitive strategies that support effective behaviors (e.g., by providing schemata for economical memory storage [6–8] and efficient attentional allocation during search [9–11]), more recent work has begun to investigate the influence of real-world structure on how we perceive and represent objects. A rapidly burgeoning literature now indicates that positional regularities affect basic perceptual analysis both in terms of neural responses in visual cortex (e.g., by shaping tuning properties of object-selective regions) and perceptual sensitivity in psychophysical tasks (e.g., by facilitating object recognition and detection). Intriguingly, the general relevance of these effects has now been demonstrated across a range of high-level visual domains, including everyday objects, faces and bodies, words, and even social interactions between people. Drawing from both the neuroimaging and behavioral literatures, in this review we synthesize recent findings across processing levels and visual domains, and discuss how their resulting insights improve our understanding of real-world object vision.

Adaptations to Absolute Locations in Individual-Object Processing

In natural environments, many objects appear at specific locations within a scene. For example, in indoor scenes, lamps are commonly found on the ceiling, whereas carpets are found on the floor. In natural vision these typical locations within a scene (in world-centered coordinates) translate to typical absolute locations within the visual field (in retinotopic coordinates). As a consequence, as we sample visual information from the scene, many objects – until directly fixated – are projected to specific locations in the visual field. Owing to their typical within-scene locations, for example, lamps tend to occur in the upper visual field and carpets tend to occur in the lower visual field (Figure 1A).

[1]Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands
[3]Berlin School of Mind and Brain, Humboldt-Universität Berlin, Berlin, Germany
[4]Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany

*Correspondence:
danielkaiser.net@gmail.com (D. Kaiser) and m.peelen@donders.ru.nl (M.V. Peelen).

Recent studies have shown that these typical absolute locations in the visual field directly influence object perception: that is, the brain processes the same object differently depending on whether it appears at its typical visual field location or elsewhere. A recent fMRI study [12] used **multivariate pattern analysis** (MVPA; see Glossary) to decode the neural representations of individual objects (e.g., lamp or carpet) presented in either their typical or atypical visual field locations. Within object-selective lateral occipital cortex (LOC), decoding was more accurate when objects appeared at their typical location (e.g., a lamp in the upper visual field) than when these same objects appeared at an atypical location (e.g., a lamp in the lower visual field). This finding suggests that regularities in the absolute location of an object affect how it is encoded in the visual system, with sharper and more discriminable representations at retinal locations that correspond to its typical location in space.

Such effects are not confined to everyday objects but also extend to other stimulus classes: in occipitotemporal cortex, individual face and body parts evoke more distinct response patterns when they appear in their typical visual field locations (e.g., an eye in the upper visual field) compared to atypical locations (e.g., an eye in the lower visual field) [13,14]. Notably, these and other studies [15] have reported behavioral recognition advantages when faces, face parts, and body parts are shown in their typical locations, suggesting that adherence to real-world spatial structure facilitates both cortical processing and perceptual performance.

Performance benefits for typically positioned objects are even observable in simple detection tasks. In **continuous flash suppression** (CFS), interocular suppression renders a visual stimulus invisible for several seconds before it can be detected, and the time until it becomes visible (i.e., 'breaks' suppression) is considered to be a sensitive measure of the detectability of an object [16,17]. In such CFS designs, objects [18] and face parts [19] break suppression faster when presented in their typical absolute locations compared to atypical locations (Figure 1B). These effects are observed even when object identity is irrelevant for the task of the participant, suggesting that basic perceptual sensitivity for high-level stimuli is increased at their typical real-world locations.

That the typical absolute location of an object can impact upon its basic perceptual processing prompts the interpretation that these effects reflect changes in neural tuning properties. Evidence supporting this interpretation comes from electrophysiological studies which show that object representations are modulated by the position of the object in the visual field very soon following stimulus onset. Within the first 140 ms of vision, representations of both objects [20] and face parts [21] are strongest when the stimuli appear in their typical absolute locations, suggesting that location biases reflect neural tuning during perceptual stimulus analysis rather than solely post-perceptual feedback.

If the effects of typical positioning do not reflect post-perceptual feedback, how are they implemented within the visual architecture? One possible explanation comes from research exploring the **receptive field** (RF) organization of category-selective regions of occipitotemporal cortex. Studies using **population receptive-field mapping** [22,23] have revealed a startling functional correspondence between RF organization and category selectivity across high-level vision, showing that the RF properties of different category-selective regions are biased towards those parts of the visual field that are typically occupied by the preferred categories of the regions (Figure 1C).

For instance, RFs in word-selective cortex of English speakers are comparably small, biased towards foveal vision, and extend further horizontally than they do vertically [24,25]. This RF

Glossary

Continuous flash suppression (CFS): a psychophysical method to study access to visual awareness. During CFS, a static stimulus is shown to one eye (e.g., via anaglyph glasses) while a dynamic contrast-rich mask is repeatedly flashed to the other eye (e.g., 10 masks per s), rendering the static stimulus invisible or 'suppressed' for a sustained period (e.g., a few seconds). The elapsed time before an observer successfully detects or localizes the stimulus is often taken as a measure of the ability of the stimulus to access visual awareness [16,17].
Gestalt: a meaningful whole emerging from the arrangements of multiple and simpler components. Gestalt principles refer to the laws that govern the integration of simple visual elements into a coherent percept.
Integrative processing: processing that combines multiple representations into composite representations.
Multivariate pattern analysis (MVPA): an analysis technique that capitalizes on pattern information in neural recordings. Whereas classical univariate analyses compare mean activations across conditions, multivariate analyses compare distributed activations. Multivariate analysis approaches are widely used to study cortical representations using fMRI [97] and magnetoencephalography/electroencephalography (M/EEG) [98].
Multivoxel combination analysis: a variant of MVPA for fMRI, where response patterns evoked by individual objects are used to model response patterns evoked by multiple objects. If a linear combination of the individual response patterns (e.g., their mean) is an accurate approximation of the group pattern, the objects are processed independently. Conversely, if the group pattern is less-accurately approximated by the linear combination of individual response patterns, additional integrative processes are involved [63].
Population receptive-field mapping: a method for estimating the RF properties of neural assemblies based on fMRI responses. Typically, high-contrast stimuli are moved across the visual field in a systematic way, and the resulting data are used to model RF positions and sizes across visual space (reviewed in [23]).
Receptive field (RF): the region of (visual) space where stimulation changes the firing behavior of a neuron.

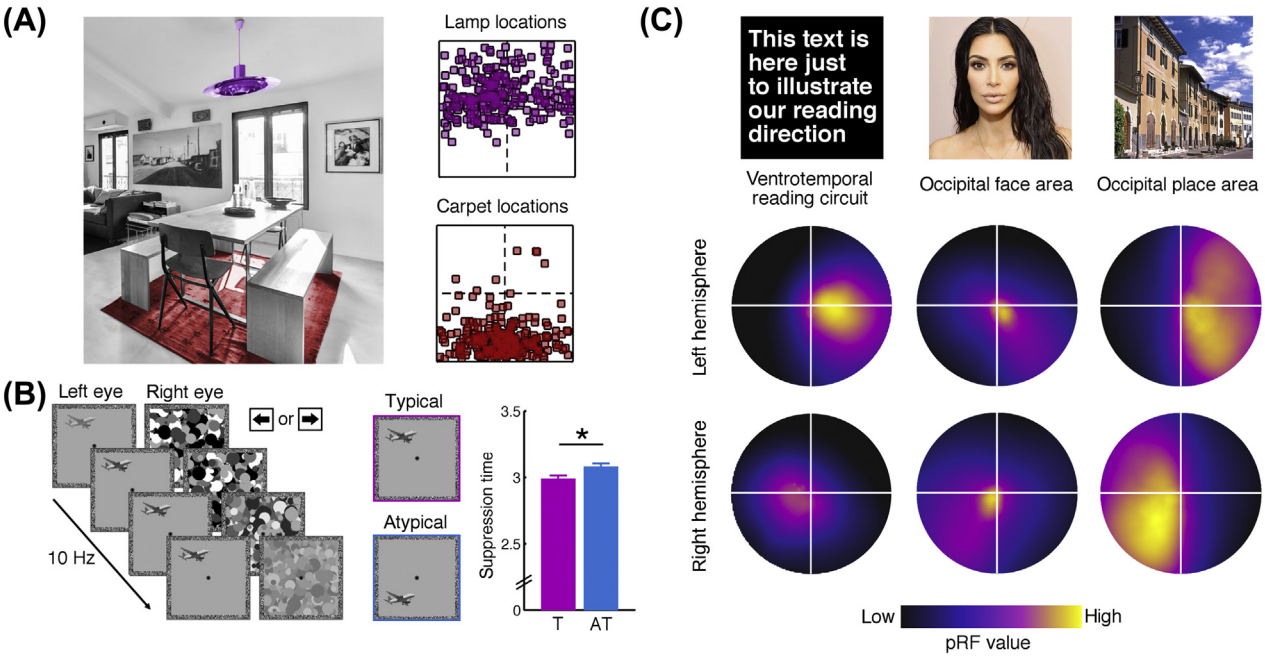

Figure 1. Neural Adaptations to Typical Absolute Object Locations. (A) The structure of natural scenes yields statistical regularities in the absolute positions of objects across visual space. Consequently, some objects tend to occupy particular visual field locations: unless directly fixated, lamps and carpets commonly appear in the upper and lower visual field, respectively. Scatter plots illustrate their position across 250 photographs from the LabelMe toolbox [129]. (B) In continuous flash suppression (CFS) experiments, the same objects gain preferential access to awareness (i.e., are detected faster) when they are typically positioned. Notably, because participants only needed to localize the target (no explicit recognition was required), the results indicate that typical absolute locations facilitate basic perceptual processing. Data reproduced from both experiments in [18]. Abbreviations: AT, atypical; T, typical. (C) Enhanced processing of typically positioned objects may be mediated by spatial neural tuning. Such tuning becomes apparent in the receptive field (RF) organization of category-selective regions, as uncovered by population RF (pRF) mapping studies. Even when measured with meaningless checkerboard stimuli, the coverage of visual space by the regions is consistent with spatial sampling of their preferred high-level contents: word-selective regions show a bias towards central vision and the horizontal meridian, face-selective regions have RFs close to the center of gaze, and scene-selective regions extensively cover peripheral space. Similarly, neurons coding individual objects may have RFs that preferentially cover areas of visual space in which these objects typically appear. Data reproduced from [24,31]. pRF maps cover 30° (words) or 20° (faces/scenes) of visual space; the color range has been adjusted to the maximum value of each region.

architecture mirrors the spatial sampling of written text during reading, which strongly relies on foveating and where information unfolds along the horizontal dimension. A complementary study [26] observed stronger responses to letters (but not to false fonts) along the horizontal meridian, corroborating the notion that word-specific activations are shaped by the direction of processing during reading.

A similar link between RF position and content-specific visual field biases is found in face- and place-selective cortices: face-selective regions have small RFs close to the center of gaze, consistent with the foveal processing necessary for individuating faces [27–31]. By contrast, place-selective regions have larger RFs that extensively cover peripheral visual space, consistent with the coarser spatial processing of natural scenes [29–32].

Taken together, these studies suggest that RF properties of high-level visual cortex are tightly linked to the characteristic spatial distribution of visual objects. Importantly, since these studies typically use meaningless checkerboard stimuli to map RF properties, their findings demonstrate that visual field biases exhibited by category-selective regions are evident even in the absence of any categorical processing demands. That is, when no categorical information is present in the stimulus, the field of view of a region cannot be adjusted based on content-specific feedback

processes. These population RF mapping studies therefore corroborate the notion that our extensive experience with real-world environments influences neural tuning independently from top-down feedback (Box 1).

The conjoint tuning to object category and visual field location yields measurable benefits in perceptual performance: across individuals, RF sizes in face-selective and word-selective cortex, respectively, predict face recognition performance [33] and reading speed [24]. At a finer-grained level, the characteristic spatial coverage of object-selective neurons may predispose the enhanced representation of typically positioned objects even within a category [12–14].

Which level of representation is enhanced when objects are positioned in their typical real-world locations? The fact that the effects of typical positioning are observed in high-level visual cortex suggests that they are not caused by visual field biases in low-level feature processing. However, these regions represent a multitude of object properties ranging from object-associated mid-level attributes (e.g., the characteristic shape or texture of an object) to categorical object content. Because these organizations are spatially entwined [34,35], it is currently unclear whether the preferential processing of typically positioned objects reflects differences in object-level representations, or in the representation of object-associated mid-level features, or both.

To summarize, recent findings provide convergent evidence that the cortical object-processing architecture is tailored to the spatial distribution of objects in the real world. Consequently, object perception varies systematically across the visual field, with more efficient processing for individual objects appearing in their typical absolute locations in the world.

Adaptations to Relative Locations in Multiobject Processing

Natural environments are inherently structured not only in terms of the absolute locations of objects within the environment, but also in terms of the relative positioning of objects with respect to each other. For example, objects in a dining room typically appear in specific relative locations (e.g., chairs typically surround a table, with a lamp above and a carpet below) (Figure 2A). Such statistical regularities in the relative positions of objects influence object processing in systematic ways, in the same way as regularities in the absolute locations of objects influence such processing.

Exactly as the typical absolute positioning of objects impacts on basic levels of perceptual processing, so too does their typical relative positioning: under CFS, observers detect groups of typically arranged objects (e.g., a lamp above a table) faster than groups of atypically arranged objects (e.g., a lamp below a table) [36] (Figure 2B), even when the task does not require explicit object recognition. Importantly, a control experiment dissociated the relative-position benefit from

Box 1. Origins of Cortical Adaptations to Real-World Structure

When and how do cortical adaptations to real-world structure emerge? One possibility is that these adaptations reflect experience-based changes in neural tuning. This view is supported by perceptual learning studies that show that cortical tuning to specific low-level features and their conjunctions is enhanced in a spatially specific way [99]. In line with this idea, recent fMRI results show that RF biases in face- and word-selective cortex are shaped across development [100,101], suggesting a key role for visual experience in the formation of RF properties in these regions. Alternatively, visual field biases could be an inherent property of the cortical architecture, and thus be in place even before visual experience plays out [30,102,103]. On this possibility, suitable neural assemblies are subsequently 'conquered' by stimuli that require their specific tuning properties. This view is supported by the observation that category-selective regions are characterized by unique structural fingerprints such as their cytoarchitecture [104,105] and connectivity with other brain regions [106,107]. Interestingly, the connectivity patterns of visual regions can be in place before experience can sculpt their functional profile [108], and are remarkably similar in the absence of visual experience [109]. In the end, both mechanisms may be at work, with pre-existing and rigid architectural properties being refined by moderate changes in cortical tuning in response to visual experience [110].

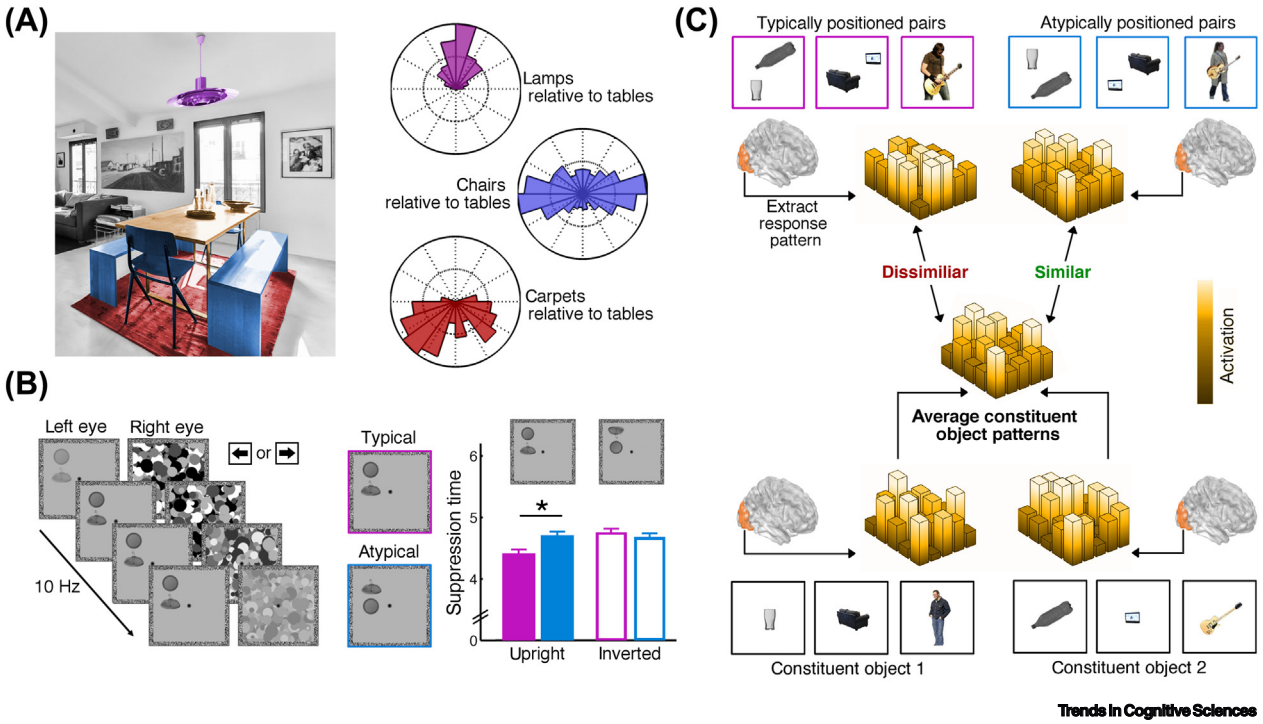

Figure 2. Neural Adaptations to Typical Relative Object Locations. (A) In addition to regularities in absolute object positions, the structure of natural scenes also yields regularities in the locations of objects relative to each other. For example, dining tables typically appear beneath lamps, above carpets, and surrounded by chairs. Polar plots illustrate the position of lamps, carpets, and chairs, all relative to tables, across photographs from the LabelMe toolbox [129]. (B) When multiple objects are positioned in their typical relative locations, they are preferentially detected under continuous flash suppression (CFS). Similarly to the effects of absolute positioning, regularities in relative positions thus grant benefits in basic perceptual processing. Notably, because stimulus inversion abolishes these effects, they are not explicable by low-level factors. Data reproduced from both experiments in [36]. (C) At a neural level, the advantages for typically positioned multiobject arrangements may arise from aggregating individual objects into group representations, as indicated by fMRI studies comparing multivoxel response patterns evoked by multiobject displays and their constituent individual objects. These studies show that multiobject response patterns are well predicted by an average of the individual-object response patterns when the objects are atypically positioned, indicating independent processing. Crucially, when the objects are typically positioned, the multiobject pattern is not as accurately predicted, indicating additional integrative neural processes. Such results have been demonstrated for grouping based on action relationships (e.g., a bottle pouring water into a glass [59]), real-world co-occurrence (e.g., a sofa facing a TV [61]), and person–object interactions (e.g., a person playing a guitar [60]).

the absolute positions of the constituent objects [36], showing that both typical relative positioning and typical absolute positions facilitate object detection under CFS.

Beyond basic detection, perceptual benefits associated with typical relative positioning are found in explicit identification and recognition tasks, where typically positioned groups of objects [37–41] and interacting groups of people [42–44] are easier to perceive. Typical relative positioning also facilitates memory: perceptual detail of multiobject and multiperson displays is more accurately maintained in visual memory when the display is arranged in accordance with real-world positional regularities [44–50], suggesting that typical relative positioning facilitates the representation of multiobject information in both perceptual and cognitive systems.

Why are typically positioned object arrangements represented more efficiently? One possibility is that multiple objects arranged in their typical relative positions are represented as a group rather than as multiple individual objects, thereby reducing the descriptive complexity of multiobject representations. For instance, a table flanked by chairs with a lamp above and carpet below may be represented as a single 'dining group', rather than as multiple individual objects. This idea is reminiscent of the study of grouping in low-level vision (Box 2), where the emergence of perceptual **Gestalt** has been associated with the grouping of different pieces of visual

Box 2. Positional Regularities in Low- and High-Level Vision

Natural environments are structured not only in terms of high-level object content but also with respect to low-level visual attributes [111,112]. A prime example in this regard is the emergence of perceptual Gestalt from the grouping of multiple simple stimuli (Figure IA). Such low-level grouping modulates neural responses both in early visual cortex and in higher-level shape-selective regions [113–115]. As for real-world object regularities, multivoxel combination analysis suggests that these activation differences reflect the integration of stimulus components into group representations [116]. Moreover, this neural integration confers a behavioral advantage in capacity-limited tasks such as visual search [117–119] or working memory [120–122]. Such analogies raise the question of the extent to which the underlying cortical adaptations to low- and high-level regularities differ. First, the two effects may be situated on different levels of the processing hierarchy because recent evidence indicates that multiobject grouping effects arise later in the hierarchy than basic object sensitivity [61]. Second, the two types of regularities may arise from different mechanisms because low- and high-level effects can be dissociated using inversion effects [36,43,48,74] (Figure IB). Despite these dissociations, the cortical tuning to low- and high-level regularities may be based on similar principles, such as the grouping of multiple elements to pass capacity bottlenecks. Future research therefore needs to explicitly juxtapose regularities on different levels to extract such common principles in their implementation.

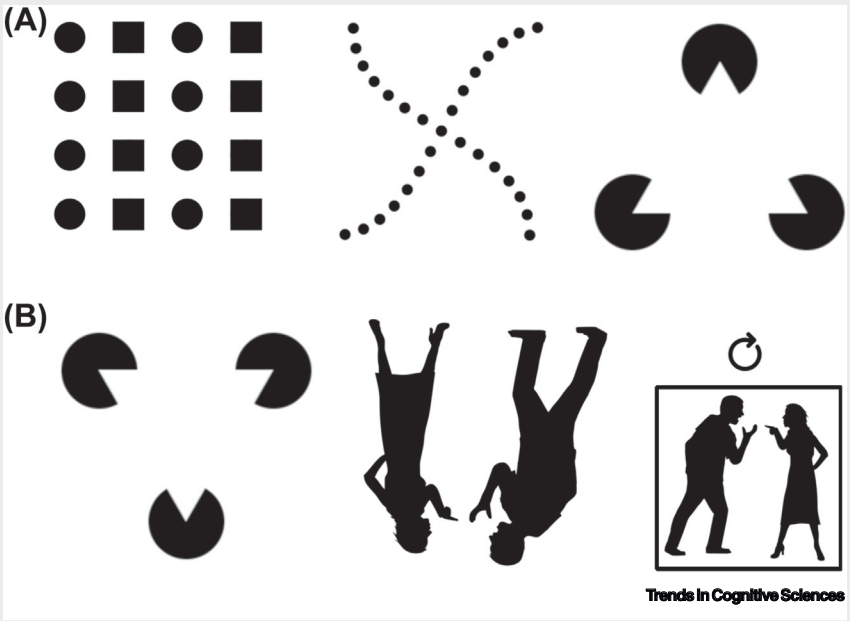

Figure I. Regularities in Low-Level Vision. (A) Examples of Gestalt formation through low-level grouping based on (left to right) similarity, good continuation, and illusory contour formation. (B) Regularities in low-level vision (e.g., grouping by contour formation) are unaffected by inversion. By contrast, high-level regularities (e.g., the grouping of multiple social agents [43]) are disrupted upon inversion.

information. Interpreted in a similar way, the studies reviewed above could reflect the grouping of objects when they appear in accordance with real-world structure.

This assertion has been tested at the neural level, where grouping is mediated by **integrative processing** of objects. This would lead to enhanced activations in visual cortex for objects in typical versus atypical relative positions. Such enhanced activity has indeed been observed for objects that co-occur in real-world scenes [51], objects that form relationships based on motor actions [52–54], faces on top of bodies [55,56], and even for multiple people engaged in social interactions [57,58].

Although these studies are in line with integrative processing, increases in overall activity may partly reflect other factors such as greater attentional engagement with the typically positioned objects. Therefore, as an alternative measure of integrative processing, recent studies have

investigated how relative object positioning affects the similarity of multivoxel response patterns in the absence of overall activity differences [59–61] (Figure 2C). These studies were inspired by the integration of simple visual features based on Gestalt laws, where 'the whole is something else than the sum of its parts' [62]. The use of **multivoxel combination analysis** [63] allows testing of whether a similar principle underlies the representation of multiple objects in visual cortex. When multiple objects are processed independently, a linear combination of the individual-object patterns (e.g., the mean) accurately predicts the multiobject pattern [64–66]. However, when multiple objects form a coherent group, the multiobject pattern (the 'whole') is relatively dissimilar to the linear combination of individual-object patterns (the 'parts').

In one such study [61], pairs of objects were positioned either as they would typically appear in real-world scenes (e.g., a sofa facing a TV) or were atypically arranged (e.g., a sofa facing away from a TV). Response patterns to the object pairs were then modeled as the mean of response patterns evoked by the constituent objects individually (e.g., sofa and TV, each in isolation). Multiobject patterns in object-selective LOC were less accurately modeled by the individual-object patterns when the objects adhered to their typical real-world positioning, providing evidence for integrative processing based on typical relative object position. Evidence for neural integration of typically positioned arrangements has also been found for other types of high-level content: for meaningful human–object interactions (e.g., a person playing a guitar), individual-object patterns did not accurately explain response patterns in the posterior superior temporal sulcus [60], and, for action relationships between objects, combination weights in LOC were altered when objects were positioned correctly for action (e.g., a bottle pouring water into a glass) [59].

At a mechanistic level, these effects are parsimoniously explained by the involvement of additional neural assemblies that exclusively represent typically positioned object groups. Through extensive exposure to concurrent objects appearing in typical relative locations, specialized neural assemblies may become tuned to the concerted presence of these objects [67–69]: consequently, these neural assemblies start to respond exclusively to the presence of the objects in their typical relative locations. These additional responses would not only enhance activations to typically positioned object groups but also distort the multivoxel response patterns they evoke. As a complementary mechanism, multiple objects may be bound by connectivity between the individual-object representations, establishing enhanced functional coupling between these representations (e.g., through neural synchrony [70,71]). Although there is evidence for such increased functional coupling between representations of features belonging to the same object [72,73], future studies need to test whether representations of multiple distinct objects can be bound in similar ways.

Another open question concerns the level of representation at which object information is grouped. Given that the effects of typical relative positions emerge in anterior parts of LOC [61], it is possible that grouping reflects an integration of high-level object representations. Alternatively, it could also reflect the integration of object-associated mid-level features, such as characteristic object shape. For example, along the vertical axis, large square-shaped objects are more often found below than above smaller objects of various forms. However, previous studies suggest that grouping does not exclusively rest on the combination of such characteristic mid-level features: grouping effects are stronger when typically positioned objects are strongly semantically related (e.g., lamp above table, mirror above sink) compared to when they are less related (e.g., mirror above table, lamp above sink) [48,74]. Similarly, grouping is reduced for people performing actions directed towards unrelated objects rather than to meaningful recipients [42].

Together, these findings suggest that the perceptual benefits observed for typically positioned multiobject arrangements reflect the grouping of multiple individual object representations. This grouping mechanism may be of particular relevance in the context of complex real-world scenes, wherein individual objects often form meaningful arrangements, such that integrating their individual representations into a higher-order group representation may serve to effectively simplify scene analysis.

Adaptations to Real-World Structure Reduce Multiobject Competition

The findings reviewed here collectively suggest that sensitivities to real-world spatial structure are ubiquitous in high-level vision: we can observe them in the neural representations of both individual objects and multiobject arrangements, as well as across a diverse range of high-level visual stimuli. We posit here that a common purpose underlies these various adaptations to real-world structure: namely, the optimal use of limited cortical resources. In the following we first reflect on the nature of cortical resource limitations, and then outline how adaptations to both typical absolute locations and typical relative locations allow us to efficiently represent objects in the context of these limitations.

A unifying commonality across perceptual and cognitive systems is their restricted capacity to process multiple entities simultaneously [75,76]. Indeed, it is well established in the low-level visual processing literature that perceptual performance is drastically compromised when multiple items compete for simultaneous representation (e.g., when searching in visual clutter [77,78]). Such difficulties in perceiving multiple objects simultaneously are tightly linked to competition effects at the neural level [79–81]: when stimuli directly compete for overlapping processing resources (e.g., when multiple objects fall within the RF of a given neuron), the response to each individual stimulus is reduced – a detrimental effect that increases in proportion with processing overlap.

In the context of real-world vision, the representational deficit imparted by interobject competition yields pessimistic predictions: most natural scenes comprise a large number of objects [82], many of which share visual and/or conceptual properties and therefore compete for the same neural processing resources. Recognizing individual objects in the face of such intense competition should be extremely challenging for the brain – but our experience in natural vision is exactly the opposite. We seem to effortlessly recognize objects even when these objects are embedded in highly complex scenes [83,84]. We suggest here that this striking discrepancy is partially accounted for by perceptual adaptations to real-world structure. Specifically, we propose that the visual system exploits the systematic spatial distribution of objects in its environment to reduce the degree to which these individual objects compete for neural processing resources. Adaptations to both the typical absolute location (in space) of an object and its typical relative location (to other objects) contribute towards this goal of reducing interobject competition. We argue that (i) typical absolute locations reduce competition through sharpened and more efficient representations of individual objects, and (ii) typical relative locations reduce competition by integrating multiple objects into group representations.

First, adaptations to the typical absolute locations of objects can reduce interobject competition by increasing the precision of neural representations. Classical theories of object recognition [85,86] typically assume that high-level vision converges towards invariant object representations which are tolerant to variation in the location of an object. Consistent with this notion, receptive fields of individual neurons in high-level visual cortex span larger areas of visual space [87]. However, at a population level, object-selective regions of the ventral visual cortex nonetheless retain relatively precise information about object location [88–90], suggesting that object recognition may not be ultimately position-invariant [91]. Indeed, to efficiently process objects without losing

information about their visual field locations, an ideal visual system would be able to support a precise representation for any given object at any possible location. In light of its limited processing resources, however, our visual system must make compromises in representational precision: because lamps reliably appear in the upper visual field, supporting a precise representation of lamps in this part of the visual field is a reasonable investment, whereas maintaining an equally precise representation in other parts of the visual field is not. As outlined above, there is mounting evidence in the neuroimaging literature for such location-governed tradeoffs in representational precision [12–14,20].

The preferential processing of particular objects by separate, spatially tuned neural populations is further apparent at the level of visual categories. In high-level visual cortex, spatially distinct regions that process information for various visual categories (e.g., scenes, faces, or words) can be differentiated both in terms of how they sample visual space [24,31] and in terms of

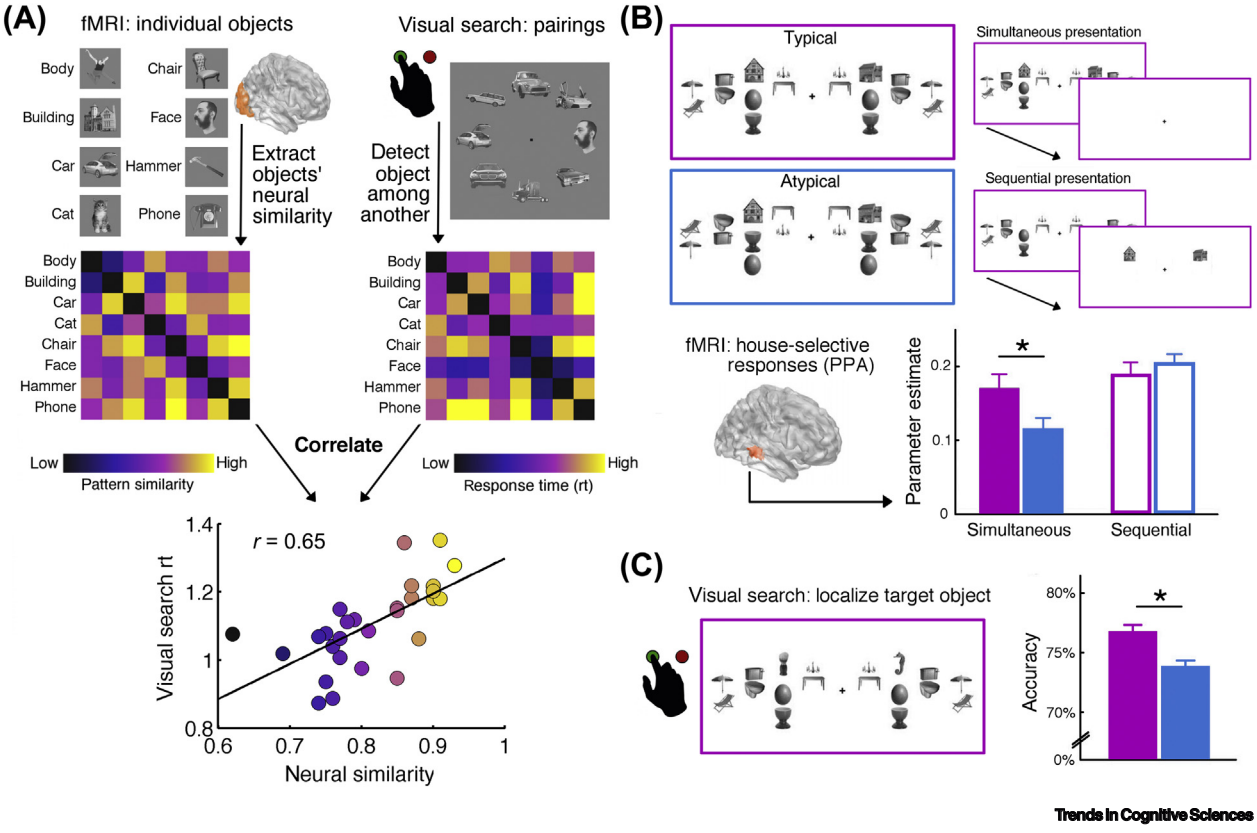

Figure 3. Adaptations to Positional Object Regularities Reduce Multiobject Competition. (A) Representational separation reduces multiobject competition. At a categorical level, visual search performance is predicted by the overlap in cortical processing of the target and distracters. For example, where there is high overlap in the neural representations of cars and telephones (i.e., they evoke similar fMRI response patterns), there is comparatively less representational overlap between cars and faces. Consequently, finding a phone among cars is comparatively slower than finding a face among cars. At a more fine-grained level, representing individual objects via distinct categorically and spatially tuned neural channels may also serve to reduce response overlap and thereby facilitate multiobject representation. Data reproduced from [95]; in the original study, stimuli were additionally matched for spatial frequency content. (B) Reducing multiobject competition by grouping objects in typical relative positions. In fMRI, unrelated objects (houses) evoke stronger selective cortical responses when surrounding object pairs conform to their typical relative positions (e.g., mirror above sink). This processing benefit for typically positioned object pairs is eliminated by temporally separating the houses and object pairs (i.e., sequential presentation), suggesting that the effect reflects reduced cortical competition between concurrent stimuli – even though neither the houses nor the object pairs were task-relevant. Abbreviation: PPA, parahippocampal place area. (C) Complementary effects are found in visual search among similar displays: consistent with the reduction of competition at a neural level, participants localize unrelated target items more accurately when distracter pairs are positioned typically rather than atypically. These results show that grouping based on typical relative positions reduces multiobject competition, thereby simplifying the perception of cluttered scenes. Data reproduced from [74].

their connections with retinotopic mechanisms at lower levels of the visual hierarchy [92]. This separation of category processing into discrete categorically and spatially tuned channels can be linked to efficiency in multiobject processing [93–95]. For example, visual search performance is determined by the cortical similarity between the target and distracter categories [95] (Figure 3A): detecting a phone among cars is difficult (because their neural representations overlap substantially and thus compete substantially), whereas detecting a face among cars is easy (because their neural representations overlap less and thus compete less). This link between processing overlap and perceptual efficiency suggests that more precise and less overlapping representations of individual objects appearing in typical locations also exhibit less competition. This mechanism may be highly beneficial in cluttered scenes that contain multiple objects, many of which appear in typical locations within the scene.

Second, adaptations to the typical relative locations of objects may reduce interobject competition by effectively reducing the number of objects competing for resources. Where a visual system with infinite processing resources could afford to process all objects in parallel (e.g., a table and a lamp), the biological constraints on the human visual brain are such that grouping objects (e.g., into a lamp-and-table group) becomes an efficient way to reduce the number of individual items competing for representation. Although the representations of object groups are still subject to resource-capacity limitations, competing for resources at a group level results in fewer representations in direct competition with one another, and consequently this competition is less detrimental than when representing the objects individually.

The notion that integrating information carried by co-occurring objects reduces interobject competition is borne out by neuroimaging work showing that object-category responses (e.g., neural activity in parahippocampal cortex evoked by a house) are stronger when concurrently presented competitor objects can be grouped based on their typical positioning (e.g., a lamp above a table, a mirror above a sink) than when they cannot be grouped based on their positioning (e.g., a table above a lamp, a sink above a mirror) [74] (Figure 3B). Notably, in this experiment neither the houses nor the competing objects were behaviorally relevant, suggesting that multiobject grouping occurs automatically during perceptual analysis. The enhanced processing of stimuli embedded in typically positioned object arrays also plays out in human behavior: in a complementary visual search experiment (Figure 3C), participants could detect targets more accurately when distracters could be grouped based on real-world regularities [74]. Together, these results suggest that interobject grouping reduces multiobject competition, and thereby simplifies the representation of complex scenes.

In sum, adaptations to real-world structure can reduce neural competition between objects in multiple, complementary ways: (i) adaptations to typical absolute locations reduce processing overlap between representations precisely tuned for particular objects appearing at particular locations, and (ii) adaptations to typical relative locations allow multiobject grouping, thereby reducing the number of objects competing for processing. Together, these adaptations simplify the neural code for scene analysis. The resulting simplification of scene representation contributes to the efficiency of human performance in naturalistic tasks, such as visual search in scenes [9,11,83,84] or scene memory [6,96].

Concluding Remarks and Future Directions

The current review underscores the fundamental and intrinsic link between the structure of natural environments and our visual perception of the world. The adaptations reviewed here support effective object processing in the real world: by capitalizing on positional regularities, the visual brain is able to optimally represent complex multiobject scenes. To conclude, we briefly revisit

Outstanding Questions

If the various adaptations to real-world positional regularities serve a common goal of simplifying scene analysis, can we also uncover common principles in their implementation? For example, is the grouping of typically co-occurring objects (e.g., a lamp above a table) qualitatively similar to the mechanism which facilitates integrating information across social agents (e.g., multiple people involved in an argument)?

Conversely, does the sensitivity of the visual system to positional regularities differ across domains? If so, how are such differences related to the different nature of regularities forming on the basis of physical constraints (e.g., object arrangements obeying the laws of physics), social situations (e.g., multiple people interacting), and societal conventions (e.g., the composition of written text in a specific language)?

Thus far, sensitivities to real-world positional regularities have typically been investigated using 2D static images. How do these adaptations manifest under more naturalistic viewing conditions, where specific objects appear at particular depths and exhibit characteristic movement patterns across time?

Does real-world structure shape perceptual architecture in modalities other than vision? Further, can future studies on real-world regularities facilitate our understanding of multisensory integration in naturalistic scenarios?

How flexible are adaptations to real-world structure? How fast can observers adapt to novel environments whose structure differs radically from the one they typically navigate in?

Can we use insights regarding the importance of scene structure to improve the design of our everyday environments? For example, can studies on real-world regularities inspire smart and easy-to-perceive designs in architecture, civil engineering, or art?

Box 3. Real-World Structure for Computer Vision

A promising avenue in which the study of adaptations to real-world regularities might fuel future developments concerns computer vision. In recent years, deep neural network (DNN) models have approached human performance in a variety of visual tasks [123,124]. These DNNs are typically trained on large and diverse sets of natural images, which inevitably exposes them to the inherent positional structures of scenes – but DNN training can also be enhanced by explicitly emphasizing real-world structure. We highlight here three ways in which adding explicit information about real-world structure could improve DNN models of vision. First, augmenting DNN training procedures with explicit information about real-world regularities (for example using human-derived object expectations [125]) could increase DNN performance levels in tasks that strongly rely on recurring spatial regularities (e.g., assisted driving). Second, in addition to improving their task performance, augmenting DNNs with real-world structure could also make their performance more human-like. Creating computational models that accurately mirror human vision would be tremendously helpful for predicting human decisions in naturalistic tasks. For example, humans are sometimes fooled by their knowledge about real-world structure, such that, in contrast to DNNs, they miss targets that do not align with their expectations (e.g., when targets are wrongly scaled given the scene context [126]). Constructing computer models that more strongly take scene structure into account, we would be able to foresee different types of errors in real-life situations and prevent them by warning humans accordingly (e.g., by warning drivers about otherwise missed hazards). Third, although DNNs have recently become the state-of-the-art model for accurately predicting visual brain activations [127,128], by no means do they explain object processing in full. Given the importance of real-world structure for human vision, enriching DNNs with positional regularity information has the potential to further improve the concordance between DNN models and the human brain. In turn, by understanding how DNNs change when explicit information about positional regularities is included during training, we can make predictions about the mechanistic implementation of regularity information in the brain.

four key insights of our review and delineate how these insights inspire future research in object vision and beyond (see Outstanding Questions).

First, adaptations to real-world structure play a key role in the perception and representation of various types of high-level content, including diverse everyday objects (e.g., furniture, tools, landmarks), human beings (e.g., faces, bodies, and their component parts), social and functional action relationships (e.g., between people and/or objects), and written text. This not only shows that high-level vision is inseparably linked to real-world structure but also highlights that positional regularities play a key role in many everyday tasks, from action understanding to reading.

Second, adaptations to real-world structure arise in both cognitive and perceptual systems. Most interestingly, they not only influence high-level processes such as recognition and working memory, but also operate at the very early stages of visual processing, even determining how quickly we detect an object in the first place. This shows that real-world positional regularities exert a more fundamental influence than was previously thought: not only do they equip humans with cognitive strategies to explore the world in smart ways, they also support the efficient perceptual parsing of natural information.

Third, the study of positional regularities in high-level vision could advance current efforts in modeling the human visual system. The far-reaching impact of real-world structure suggests that object vision cannot be fully understood without taking real-world structure into account. The recent insights thus urge a consideration of positional regularities in neural models of object processing. Interestingly, explicitly considering real-world structure may not only help to understand the biological brain but also fuel developments in computer vision (Box 3).

Finally, the importance of real-world structure supports neurocognitive research that pushes towards more naturalistic approaches to vision: only by studying vision under conditions that more closely mimic the properties of real-world environments will we come closer to understanding how we efficiently select, recognize, and ultimately extract meaning from a complex visual world.

References

1. Gauthier, I. and Tarr, M.J. (2016) Visual object recognition: do we (finally) know more now than we did? *Annu. Rev. Vis. Sci.* 2, 377–396
2. Grill-Spector, K. and Weiner, K.S. (2014) The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15, 536–548

3. Kourtzi, Z. and Connor, C.E. (2011) Neural representations for object perception: structure, category, and adaptive coding. *Annu. Rev. Neurosci.* 34, 45–67
4. Martin, A. (2007) The representation of object concepts in the brain. *Annu. Rev. Psychol.* 58, 25–45
5. op de Beeck, H.P. *et al.* (2008) Interpreting fMRI data: maps, modules and dimensions. *Nat. Rev. Neurosci.* 9, 123–135
6. Konkle, T. *et al.* (2010) Scene memory is more detailed than you think: the role of categories in visual long-term memory. *Psychol. Sci.* 21, 1551–1556
7. Mandler, J.M. and Johnson, N.S. (1976) Some of the thousand words a picture is worth. *J. Exp. Psychol. Hum. Learn. Mem.* 2, 529–540
8. Wolfe, J.M. (1998) Scene memory: what do you know about what you saw? *Curr. Biol.* 8, 303–304
9. Peelen, M.V. and Kastner, S. (2014) Attention in the real world: toward understanding its neural basis. *Trends Cogn. Sci.* 18, 242–250
10. Torralba, A. *et al.* (2006) Contextual guidance of eye movements and attention in real-world scenes: the role of global features in objects search. *Psychol. Rev.* 113, 766–786
11. Wolfe, J.M. *et al.* (2011) Visual search in scenes involves selective and nonselective pathways. *Trends Cogn. Sci.* 15, 77–84
12. Kaiser, D. and Cichy, R.M. (2018) Typical visual-field locations enhance processing in object-selective channels of human occipital cortex. *J. Neurophysiol.* 120, 848–853
13. Chan, A.W. *et al.* (2010) Cortical representations of bodies and faces are strongest in commonly experienced configurations. *Nat. Neurosci.* 13, 417–418
14. de Haas, B. *et al.* (2016) Perception and processing of faces in the human brain is tuned to typical feature locations. *J. Neurosci.* 36, 9289–9302
15. Quek, G.L. and Finkbeiner, M. (2014) Face-sex categorization is better above fixation than below: evidence from the reach-to-touch paradigm. *Cogn. Affect. Behav. Neurosci.* 14, 1407–1419
16. Gayet, S. *et al.* (2014) Breaking continuous flash suppression: competing for consciousness on the pre-semantic battlefield. *Front. Psychol.* 5, 460
17. Stein, T. *et al.* (2011) Breaking continuous flash suppression: a new measure of unconscious processing during interocular suppression? *Front. Hum. Neurosci.* 5, 167
18. Kaiser, D. and Cichy, R.M. (2018) Typical visual-field locations facilitate access to awareness for everyday objects. *Cognition* 180, 118–122
19. Moors, P. *et al.* (2016) Faces in commonly experienced configurations enter awareness faster due to their curvature relative to fixation. *PeerJ* 4, e1565
20. Kaiser, D. *et al.* (2018) Typical retinotopic locations impact the time course of object coding. *NeuroImage* 176, 372–379
21. Issa, E.B. and DiCarlo, J.J. (2012) Precedence of the eye region in neural processing of faces. *J. Neurosci.* 32, 16666–16682
22. Dumoulin, S.O. and Wandell, B.A. (2008) Population receptive field estimates in human visual cortex. *NeuroImage* 39, 647–660
23. Wandell, B.A. and Winawer, J. (2015) Computational neuroimaging and population receptive fields. *Trends Cogn. Sci.* 19, 349–357
24. Le, R. *et al.* (2017) The field of view available to the ventral occipito-temporal reading circuitry. *J. Vis.* 17, 6
25. Wandell, B.A. and Le, R. (2017) Diagnosing the neural circuitry of reading. *Neuron* 96, 298–311
26. Chang, C.H. *et al.* (2015) Adaptation of the human visual system to the statistics of letters and line configurations. *NeuroImage* 120, 428–440
27. Grill-Spector, K. *et al.* (2017) The functional neuroanatomy of human face perception. *Annu. Rev. Vis. Sci.* 3, 167–196
28. Kay, K.N. *et al.* (2015) Attention reduces spatial uncertainty in human ventral temporal cortex. *Curr. Biol.* 25, 595–600
29. Levy, I. *et al.* (2001) Center–periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539
30. Malach, R. *et al.* (2002) The topography of high-order human object areas. *Trends Cogn. Sci.* 6, 176–184
31. Silson, E.H. *et al.* (2016) Evaluating the correspondence between face-, scene- and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J. Vis.* 16, 14
32. Silson, E.H. *et al.* (2015) A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *J. Neurosci.* 35, 11921–11935
33. Witthoft, N. *et al.* (2016) Reduced spatial integration in the ventral visual cortex underlies face recognition deficits in developmental prosopagnosia. *bioRxiv*. Published online April 29, 2016. https://doi.org/10.1101/051102
34. Bracci, S. and op de Beeck, H.P. (2016) Dissociations and associations between shape and category representations in the two visual pathways. *J. Neurosci.* 36, 432–444
35. Proklova, D. *et al.* (2016) Disentangling representations of object shape and object category in human visual cortex: the animate–inanimate distinction. *J. Cogn. Neurosci.* 28, 680–692
36. Stein, T. *et al.* (2015) Interobject grouping facilitates visual awareness. *J. Vis.* 15, 10
37. Biederman, I. *et al.* (1982) Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 14, 143–177
38. Green, C. and Hummel, J.E. (2006) Familiar interacting pairs are perceptually grouped. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 1107–1119
39. Gronau, N. and Shachar, M. (2014) Contextual integration of visual objects necessitates attention. *Atten. Percept. Psychophys.* 76, 695–714
40. Riddoch, M.J. *et al.* (2003) Seeing the action: neuropsychological evidence for action-based effects on object selection. *Nat. Neurosci.* 6, 82–89
41. Roberts, K.L. and Humphreys, G.W. (2011) Action relationships facilitate the identification of briefly-presented objects. *Atten. Percept. Psychophys.* 73, 597–612
42. Papeo, L. and Abassi, E. (2019) Seeing social events: the visual specialization for dyadic human–human interactions. *J. Exp. Psychol. Hum. Percept. Perform.* Published online April 18, 2019. https://doi.org/10.1037/xhp0000064
43. Papeo, L. *et al.* (2017) The two-body inversion effect. *Psychol. Sci.* 28, 369–379
44. Vestner, T. *et al.* (2019) Bound together: social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *J. Exp. Psychol. Gen.* Published online January 17, 2019. http://dx.doi.org/10.1037/xge0000545
45. Ding, X. *et al.* (2017) Two equals one: two human actions during social interaction are grouped as one unit in working memory. *Psychol. Sci.* 28, 1311–1320
46. Draschkow, D. and Võ, M.L-H. (2017) Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Sci. Rep.* 7, 16471
47. Gronau, N. and Shachar, M. (2015) Contextual consistency facilitates long-term memory of perceptual detail in barely seen images. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1095–1111
48. Kaiser, D. *et al.* (2015) Real-world spatial regularities affect visual working memory for objects. *Psychon. Bull. Rev.* 22, 1784–1790
49. O'Donnell, R.E. *et al.* (2018) Semantic and functional relationships among objects increase the capacity of visual working memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 44, 1151–1158
50. Tibon, R. *et al.* (2014) Associative recognition processes are modulated by the semantic unitizability of memoranda. *Brain Cogn.* 92, 19–31
51. Kim, J.G. and Biederman, I. (2011) Where do objects become scenes? *Cereb. Cortex* 21, 1738–1746
52. Gronau, N. *et al.* (2008) Integrated contextual representation for objects' identities and their locations. *J. Cogn. Neurosci.* 20, 371–388
53. Kim, J.G. *et al.* (2011) The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the intraparietal sulcus: a transcranial magnetic stimulation study. *J. Neurosci.* 31, 8320–8324

54. Roberts, K.L. and Humphreys, G.W. (2010) Action relationships concatenate representations of separate objects in the ventral visual system. *NeuroImage* 52, 1541–1548
55. Bernstein, M. *et al.* (2014) An integrated face-body representation in the fusiform gyrus but not the lateral occipital cortex. *J. Cogn. Neurosci.* 26, 2469–2478
56. Song, Y. *et al.* (2013) Representation of contextually related multiple objects in the human ventral visual pathway. *J. Cogn. Neurosci.* 25, 1261–1269
57. Quadflieg, S. *et al.* (2015) The neural basis of perceiving person interactions. *Cortex* 70, 5–20
58. Walbrin, J. *et al.* (2018) Neural responses to visually observed social interactions. *Neuropsychologia* 112, 31–39
59. Baeck, A. *et al.* (2013) The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *NeuroImage* 70, 37–47
60. Baldassano, C. *et al.* (2017) Human–object interactions are more than the sum of their parts. *Cereb. Cortex* 27, 2276–2288
61. Kaiser, D. and Peelen, M.V. (2018) Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage* 169, 334–341
62. Koffka, K. (Ed.), 1935.. Principles of Gestalt Psychology. Harcourt Brace
63. Kubilius, J. *et al.* (2015) Brain-decoding reveals how wholes relate to the sum of parts. *Cortex* 72, 5–14
64. MacEvoy, S.P. and Epstein, R.A. (2009) Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Curr. Biol.* 19, 943–947
65. Kaiser, D. *et al.* (2014) Whole person-evoked fMRI activity patterns in human fusiform gyrus are accurately modeled by a linear combination of face- and body-evoked activity patterns. *J. Neurophysiol.* 111, 82–90
66. Reddy, L. *et al.* (2009) Attention and biased competition in multi-voxel object representations. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21447–21452
67. Baker, C.I. *et al.* (2002) Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. *Nat. Neurosci.* 5, 1210–1216
68. Messinger, A. *et al.* (2001) Neural representations of stimulus associations develop in the temporal lobe during learning. *Proc. Natl. Acad. Sci. U. S. A.* 98, 12239–12244
69. Sakai, K. and Miyashita, Y. (1991) Neural organization for the long-term memory of paired associates. *Nature* 354, 152–155
70. Hummel, J.E. and Biederman, I. (1992) Dynamic binding in a neural network for shape recognition. *Psychol. Rev.* 99, 480–517
71. Singer, W. (1999) Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24, 49–65
72. Gray, C.M. *et al.* (1989) Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature* 338, 334–337
73. Martin, A.B. and von der Heydt, R. (2015) Spike synchrony reveals emergence of proto-objects in visual cortex. *J. Neurosci.* 35, 6860–6870
74. Kaiser, D. *et al.* (2014) Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 111, 11217–11222
75. Broadbent, D. (Ed.), 1958.. Perception and Communication. Pergamon Press
76. Franconeri, S.L. *et al.* (2013) Flexible cognitive resources: competitive content maps for attention and memory. *Trends Cogn. Sci.* 17, 134–141
77. Treisman, A.M. and Gelade, G. (1980) A feature-integration theory of attention. *Cogn. Psychol.* 12, 97–136
78. Wolfe, J.M. *et al.* (1989) Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 419–433
79. Desimone, R. and Duncan, J. (1995) Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 18, 193–222
80. Kastner, S. and Ungerleider, L.G. (2001) The neural basis of biased competition in human visual cortex. *Neuropsychologia* 39, 1263–1276
81. Miller, E.K. *et al.* (1993) Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Res.* 616, 25–29
82. Wolfe, J.M. *et al.* (2011) Visual search for arbitrary objects in real scenes. *Atten. Percept. Psychophys.* 73, 1650–1671
83. Li, F.F. *et al.* (2002) Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U. S. A.* 99, 9596–9601
84. Thorpe, S. *et al.* (1996) Speed of processing in the human visual system. *Nature* 381, 520–522
85. Riesenhuber, M. and Poggio, T. (1999) Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 2, 1019–1025
86. DiCarlo, J.J. and Cox, D.D. (2007) Untangling invariant object recognition. *Trends Cogn. Sci.* 11, 333–341
87. Kravitz, D.J. *et al.* (2013) The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* 17, 26–49
88. Cichy, R.M. *et al.* (2011) Encoding the identity and location of objects in human LOC. *NeuroImage* 54, 2297–2307
89. Golomb, J.D. and Kanwisher, N. (2012) Higher level visual cortex represents retinotopic, not spatiotopic, object location. *Cereb. Cortex* 22, 2794–2810
90. Hemond, C.C. *et al.* (2007) A preference for contralateral stimuli in human object- and face-selective cortex. *PLoS One* 2, e574
91. Kravitz, D.J. *et al.* (2008) How position dependent is visual object recognition? *Trends Cogn. Sci.* 12, 114–122
92. Uyar, F. *et al.* (2016) Retinotopic information interacts with category selectivity in human ventral cortex. *Neuropsychologia* 92, 90–106
93. Cohen, M.A. *et al.* (2014) Processing multiple visual objects is limited by overlap in neural channels. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8955–8960
94. Cohen, M.A. *et al.* (2015) Visual awareness is limited by the representational architecture of the visual system. *J. Cogn. Neurosci.* 27, 2240–2252
95. Cohen, M.A. *et al.* (2017) Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *J. Neurophysiol.* 117, 388–402
96. Hollingworth, A. (2004) Constructing visual representations of natural scenes: the roles of short- and long-term visual memory. *J. Exp. Psychol. Hum. Percept. Perform.* 30, 519–537
97. Haynes, J.D. (2015) A primer on pattern-based approaches to fMRI: principles, pitfalls, and perspectives. *Neuron* 87, 257–270
98. Contini, E.W. *et al.* (2017) Decoding the time-course of object recognition in the human brain: from visual features to categorical decisions. *Neuropsychologia* 105, 165–176
99. Sasaki, Y. *et al.* (2010) Advances in visual perceptual learning and plasticity. *Nat. Rev. Neurosci.* 11, 53–60
100. Gomez, J. *et al.* (2018) Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nat. Commun.* 9, 788
101. Gomez, J. *et al.* (2018) Development of population receptive fields in the lateral visual stream improves spatial coding amid stable structural-functional coupling. *NeuroImage* 188, 59–69
102. Dehaene, S. and Cohen, L. (2007) Cultural recycling of cortical maps. *Neuron* 56, 384–398
103. Srihasam, K. *et al.* (2014) Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nat. Neurosci.* 17, 1776–1783
104. Weiner, K.S. *et al.* (2014) The mid-fusiform sulcus: a landmark identifying both cytoarchitectonic and functional divisions of human ventral temporal cortex. *NeuroImage* 84, 453–465
105. Weiner, K.S. and Zilles, K. (2016) The anatomical and functional specialization of the fusiform gyrus. *Neuropsychologia* 83, 48–62
106. Osher, D.E. *et al.* (2016) Structural connectivity fingerprints predict cortical selectivity for multiple visual categories across cortex. *Cereb. Cortex* 26, 1668–1683
107. Saygin, Z.M. *et al.* (2011) Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. *Nat. Neurosci.* 15, 321–327

108. Saygin, Z.M. *et al.* (2016) Connectivity precedes function in the development of the visual word form area. *Nat. Neurosci.* 19, 1250–1255
109. Wang, X. *et al.* (2015) How visual is the visual cortex? Comparing connectional and functional fingerprints between congenitally blind and sighted individuals. *J. Neurosci.* 35, 12545–12559
110. op de Beeck, H.P. and Baker, C.I. (2010) The neural basis of visual object learning. *Trends Cogn. Sci.* 14, 22–30
111. Geisler, W.S. (2008) Visual perception and the statistical properties of natural scenes. *Annu. Rev. Psychol.* 59, 167–192
112. Purves, D. *et al.* (2011) Understanding vision in wholly empirical terms. *Proc. Natl. Acad. Sci. U. S. A.* 108, 15588–15595
113. Altmann, C.F. *et al.* (2003) Perceptual organization of local elements into global shapes in the human visual cortex. *Curr. Biol.* 13, 342–349
114. Fang, F. *et al.* (2008) Perceptual grouping and inverse activity patterns in human visual cortex. *J. Vis.* 8, 2
115. Murray, S.O. *et al.* (2002) Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15164–15169
116. Kubilius, J. *et al.* (2011) Emergence of perceptual Gestalts in the human visual cortex: the case of the configural-superiority effect. *Psychol. Sci.* 22, 1296–1303
117. Donnelly, N. *et al.* (1991) Parallel computation of primitive shape descriptions. *J. Exp. Psychol. Hum. Percept. Perform.* 17, 561–570
118. Humphreys, G.W. *et al.* (1989) Grouping processes in visual search: effects with single- and combined-feature targets. *J. Exp. Psychol. Gen.* 118, 258–279
119. Rauschenberger, R. and Yantis, S. (2006) Perceptual encoding efficiency in visual search. *J. Exp. Psychol. Gen.* 135, 116–131
120. Brady, T.F. and Tenenbaum, J.B. (2013) A probabilistic model of visual working memory: incorporating higher order regularities into working memory capacity estimates. *Psychol. Rev.* 120, 85–109
121. Woodman, G.F. *et al.* (2003) Perceptual organization influences visual working memory. *Psychon. Bull. Rev.* 10, 80–87
122. Xu, Y. (2006) Understanding the object benefit in visual short-term memory: the roles of feature proximity and connectedness. *Percept. Psychophys.* 68, 815–828
123. Krizhevsky, A. *et al.* (2012) Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst.* 1, 1097–1105
124. LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436–444
125. Katti, H. *et al.* (2019) Machine vision benefits from human contextual expectations. *Sci. Rep.* 9, 2112
126. Eckstein, M.P. *et al.* (2017) Humans, but not deep neural networks, often miss giant targets in scenes. *Curr. Biol.* 27, 2827–2832
127. Cichy, R.M. and Kaiser, D. (2019) Deep neural networks as scientific models. *Trends Cogn. Sci.* 23, 305–317
128. Kriegeskorte, N. (2015) Deep neural networks: a new framework for modeling biological vision and brain information processing. *Annu. Rev. Vis. Sci.* 1, 417–446
129. Russell, B.C. *et al.* (2008) LabelMe: a database and web-based tool for image annotation. *Int. J. Comput. Vis.* 77, 157–173